

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/353352128>

Classifying precipitation from GEO Satellite Observations: Prognostic Model

Article in Quarterly Journal of the Royal Meteorological Society · July 2021

DOI: 10.1002/qj.4134

CITATIONS

0

READS

64

5 authors, including:



Shruti Upadhyaya

University of Oklahoma

16 PUBLICATIONS 74 CITATIONS

[SEE PROFILE](#)



Pierre-Emmanuel Kirstetter

University of Oklahoma

159 PUBLICATIONS 3,598 CITATIONS

[SEE PROFILE](#)



Robert Kuligowski

National Oceanic and Atmospheric Administration

83 PUBLICATIONS 1,722 CITATIONS

[SEE PROFILE](#)



Jonathan Gourley

NOAA's National Severe Storms Laboratory

241 PUBLICATIONS 6,641 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Precipitation Estimation from Geostationary Satellites [View project](#)



Clouds and Atmospheric Water Vapor [View project](#)

Classifying precipitation from GEO Satellite Observations: Prognostic Model

Shruti A. Upadhyaya^{1,*}, Pierre-Emmanuel Kirstetter^{2,3,4,5,*}, Robert J. Kuligowski⁶, Jonathan J. Gourley⁵, Heather Grams⁵

¹ Cooperative Institute for Mesoscale Meteorological Studies, Norman, Oklahoma

² School of Meteorology, University of Oklahoma, Norman, Oklahoma

³ School of Civil Engineering and Environmental Science, University of Oklahoma, Norman, Oklahoma

⁴ Advanced Radar Research Center, University of Oklahoma, Norman, Oklahoma

⁵ NOAA/National Severe Storms Laboratory, Norman, Oklahoma

⁶ NOAA/NESDIS/Center for Satellite Applications and Research, College Park, Maryland

*Corresponding authors: Pierre-Emmanuel Kirstetter (pierre.kirstetter@noaa.gov);

Shruti A. Upadhyaya (shruti.a.upadhyaya-1@ou.edu)

Funding Information:

Funding for this research was provided by the GOES-R Series Risk Reduction program, which provided support to the Cooperative Institute for Mesoscale Meteorological Studies at the University of Oklahoma under Grant NA16OAR4320115. P. Kirstetter acknowledges support from NASA Global Precipitation Measurement Ground Validation program under Grant NNX16AL23G and Precipitation Measurement Missions program under Grant 80NSSC19K0681

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: [10.1002/qj.4134](https://doi.org/10.1002/qj.4134)

Key Points:

- A machine learning precipitation classification model is developed for GOES-16 satellite observations and numerical weather predictions.
- Promising results in identifying the occurrence/non-occurrence with accuracy 93%
- Hail and Warm Stratiform types have high detection scores greater than 70%
- Challenges exists in separating Convective types

Abstract

Precipitation is one of the most important components of the global water and energy cycles, which together regulate the climate system. Future changes in precipitation patterns related to climate change are likely to bear the largest impacts on society. The new generation of geostationary Earth orbit (GEO) satellites provide high-resolution observations and opportunities to improve our understanding of precipitation processes. This study contributes to improved precipitation characterization and retrievals from space by identifying precipitation types (e.g., convective, stratiform) with multi-spectral observations from the Advanced Baseline Imager (ABI) sensor onboard the GOES-16 satellite. A machine learning-based classification model is developed by deriving a comprehensive set of features using five ABI channels and numerical weather prediction observations, and trained with the Ground Validation Multi-Radar/Multi-Sensor (GV-MRMS) system used as a benchmark. The developed prognostic model shows skillful performance in identifying the occurrence/non-occurrence of precipitation (accuracy of 97%; Kappa coefficient of 0.9) and precipitation processes, with an overall classification accuracy of 76% and Kappa coefficient of 0.56. Challenges exist in separating convective and tropical from other precipitation types. It is suggested to utilize probabilities instead of deterministically separating precipitation types especially in regions with uncertain classifications.

Keywords: GOES-16, Numerical Weather Prediction, Precipitation, Machine Learning, Classification, Geostationary Satellites

1. Introduction

With enhanced observation capabilities, the new generation of geostationary (GEO) satellites provides an opportunity to improve the observation and estimation from space of a major component of the water and energy cycles - precipitation. NOAA's Advanced Baseline Imager (ABI) sensor onboard the Geostationary Operational Environmental Satellites (GOES-R Series) provides three times more spectral channels, four times the resolution, and five times faster scanning across North and South America when compared to its predecessor imager onboard previous-generation GOES (Schmit et al., 2017). Likewise, other GEO sensors with more than 10 spectral channels such as the Advanced Himawari Imager (AHI) onboard Himawari 8-9, the Advanced Geosynchronous Radiation Imager (AGRI) onboard FY-4A, the Advanced Meteorological Imager (AMI) onboard GEO-KOMPSAT-2A, together with ABI (GOES-R) provide global coverage at high spatial and temporal resolution. An overarching challenge is now to effectively utilize and explore this huge information gain for science and applications. Specifically, the improved temporal, spatial and spectral resolution of precipitation observations support our understanding of highly spatially and temporally varying precipitation processes at mesoscale and thereby improving global precipitation retrievals.

Retrieving surface precipitation from GEO observations is challenging because the Visible (VIS)/Infrared (IR) regions of the electromagnetic spectrum provide mainly cloud-top related information. The relation between cloud top brightness temperatures and surface precipitation rates varies with precipitation types (e.g., convective versus stratiform precipitation) and the vertical distribution of their processes (Vicente et al., 1998). Therefore, most Quantitative Precipitation Estimation (QPE) algorithms with active or passive remote-sensing instruments initially classify the observed precipitation into different categories before applying separate

Accepted Article

observations-to-surface rate relations to improve the retrieval accuracy. This is an active research domain (e.g. active microwave observations: Le and Chandrasekar, 2012, 2021; passive microwave observations: Petković et al., 2019 ; VIS/IR: Grams et al., 2016; Thies et al., 2008). Challenges in precipitation quantification from GEO sensors arise at the initial stage of detecting precipitation processes (Upadhyaya et al., 2020).

Most QPE operational algorithms use part of the available channels from ABI. The Self-Calibrating Multivariate Precipitation Retrieval (SCaMPR; Kuligowski 2002; Kuligowski et al. 2016) uses 5 out of the 16 ABI channels, and the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks - Cloud Classification System (PERSIANN-CCS; Hong et al., 2004) only uses one channel (i.e., 11.2 μ m). While spatial features derived from these channels (9 features from PERSIANN-CCS and 2 for SCaMPR) provide additional information, several recent studies showed that more features can be derived to aid in improved precipitation detection, classification and quantification (Giannakos and Feidas, 2013; Thebbi and Haddad, 2016; Thies et al., 2008; Upadhyaya and Ramsankaran, 2014). As a result, the information derived from new channels provided by the recent generation of GEO sensors such as ABI is still underexplored, while challenges remain in detecting and quantifying precipitation types such as non-precipitating cold cirrus clouds or warm/shallow precipitation from space-based platforms (So and Shin, 2018). For the first time to our knowledge, a framework is designed to consistently and systematically analyse satellite-based indices for precipitation detection and classification.

An endemic limitation is the limited availability of standard reference precipitation typologies, which leads most operational algorithms to utilize unsupervised techniques. For example, SCaMPR and PERSIANN-CCS utilize techniques solely based on the ABI data that do

not explicitly identify precipitation systems or types (Grams et al., 2016). Their outputs allow for limited physical interpretation of precipitation types observed by ABI.

Reference data play a key role in the precipitation classification or type identification problem to interpret and extract physical information observed by GEO sensors. Most of the recent research studies either use expert classified images as reference (Tian et al., 1999; Saitwal et al., 2003) which limits the volume of data used, or gauge-based surface precipitation rates (Giannakos and Feidas, 2013; Thebbi and Haddad, 2016) to separate convective and stratiform precipitation based on precipitation intensity and duration rather than on the basis of physical processes. Active remote sensors (i.e., radars) from ground and space provide more physical depiction of precipitation processes and thus provide more accurate classifications as they elucidate cloud microphysical properties. Only a few studies have attempted to utilize this information to guide the training of classification models and understand the information from GEO observations (Grams et al., 2016; So and Shin, 2018; Thies et al., 2008). However, these studies have been carried out with few spectral channels only and with limited use of derived features or cloud-top properties.

The overall objective of this work is to effectively utilize GOES-16 ABI observations to identify precipitation types as seen by ground-based radars in order to improve precipitation retrievals. The research questions we attempt to answer are: How much information on precipitation processes or types as observed from ground radars is contained in ABI cloud-top observations? Can they be used to discriminate the precipitation processes or types identified with active sensors? How much do modeled environmental conditions complement the ABI observations?

Accepted Article

In this study, we propose to use five day-night channels from GOES-16 ABI observations (as in SCaMPR in order to facilitate direct comparison with it) to develop a comprehensive machine learning model for precipitation type classification. A comprehensive set of features are derived and tested, many of which are derived for the first time. Information about low-level environmental conditions complement the ABI cloud-top information and are derived using Numerical Weather Prediction (NWP) model-based environmental features. The reference used is the set of surface precipitation types provided by the Ground-Validation Multi-Radar/Multi-Sensor (GV-MRMS) system (Kirstetter et al. 2018) across the conterminous United States (CONUS).

Following the introduction, Section 2 introduces data and methods with detailed discussion on newly derived indices. Section 3 reports the results and analyzes the experiments carried out, and the last section concludes and summarizes takeaway points from this work.

2. Data and Methods

2.1. GEO Satellite: GOES-R Advanced Baseline Imager (ABI) and derived features

To derive and test a comprehensive set of features, five ABI channels (Channel 8: 6.2 μ m, 10: 7.3 μ m, 11: 8.5 μ m, 14: 11.2 μ m and 15: 12.3 μ m) are used in this study. The brightness temperatures are parallax-adjusted (Kuligowski et al., 2016). These channels are selected based on their availability during both day and nighttime, and they have equivalents in other GEO sensors (e.g. channels 5, 6, 7, 9 and 10 in METEOSAT Spinning Enhanced Visible InfraRed Imager), thus making this study more globally applicable to other sensors. The spatial resolution of the selected ABI channels is approximately 2 km at nadir. Several categories of features are derived from the ABI observations and from the NWP analyses (discussed in the coming section) and are listed in Table 1.

Early work by Lovejoy and Austin (1979) showed that multi-spectral data can improve upon single-channel approaches to detect and quantify precipitation. Since then, several channels and combinations of channels have been used as additional features. The first category of features listed in Table 1 includes Brightness Temperatures (BTs) from the 5 channels. The water vapor (WV) absorption channels (6.2 μ m and 7.3 μ m) are sensitive to different levels of tropospheric WV. The 6.2 μ m (7.3 μ m) band is more sensitive to upper(lower)-level tropospheric WV, while the IR window channels (8.5 μ m, 11.2 μ m and 12.3 μ m) include the 11.2 μ m band that is one of the most important heritage channels. With greater absorption due to WV than the 11.2 μ m channel, the 8.5 μ m and 12.3 μ m channels are referred to as “dirty” IR bands.

The second category of features includes Brightness Temperature Differences (BTD; see Category 2 in Table 1). BTD is the most common way of combining information from two channels and has been significantly explored in the literature (Ba and Gruber, 2001; Kuligowski et al., 2016; Tjemkes et al., 1997; Upadhyaya and Ramsankaran, 2014). For example, the difference between IR and WV channels is used to separate overshooting cloud tops and cirrus clouds (Tjemkes et al., 1997) and BTD between two IR window channels is used to detect the cloud phase (Baum and Platnick, 2006; Giannakos and Feidas, 2013).

The third category of features includes Differences of BTDs (D-BTD; Category 3 in Table 1). Recently, So and Shin (2018) used a D-BTD feature; i.e., (8.5 μ m,-11.2 μ m) - (11.2 μ m- 12.3 μ m) to improve cloud phase detection. In this study we expand on this concept and all possible combinations of BTDs and D-BTDs are derived with new channels to explore their usefulness in detecting different precipitation processes.

The fourth category of features are textures (Te; Category 4 in Table 1). Textures are the representations of spatial characteristics of a surface (Mohanaiah et al., 2013). The texture features

Accepted Article

derived from several GEO sensors channels are found to be useful at all stages of the precipitation retrieval process (Ba and Gruber, 2001; Kuligowski et al., 2016; Hong et al., 2004; Giannakos and Feidas, 2013; Tian et al., 1999). The usefulness of textures derived from all channels, difference indices (BTD) and difference of difference (D-BTD) has not yet been studied. The frequently used Grey Level Co-occurrence Matrix (GLCM; Haralick et al., 1973) for satellite images is used to derive texture features. GLCM extends texture measures beyond first order (such as mean and variance) by describing second-order features in the satellite image (Hall-Beyer, 2000). GLCM broadly estimates the joint probability density function of grey level pairs in an image (Rampun et al., 2013; Xian, 2010). In this study, grey levels are brightness temperatures, BTDs, and D-BTDs. Using the GLCM matrix across 5 x 5 ABI grids in all four orientations (0, 45, 90 and 135 degrees), five texture features are derived, namely: "mean", "variance", "entropy", "homogeneity", and "contrast". Mean and variance are the descriptive statistics. Entropy is a measure of disorder or randomness (information content). Contrast is a measure of local variations. Homogeneity is inversely proportional to contrast at constant energy (Xian, 2010). For details on the computation and exact definitions of these textures, the reader is referred to Hall-Beyer (2000). In this study, the R package 'glcm' (Zvoleff, 2020) is used to compute texture features.

To account for the impact of varying viewing geometries on the spatial resolution of the ABI sensor at the surface, satellite zenith angle is also used as a feature (Ze; Category 5 in Table 1). The last category of features is derived from NWP model analyses, which give information about low-level environmental conditions and complement the ABI cloud-top information. More details of this category are discussed in Section 2.3. In total, 260 different features derived from 5 channels and NWP 1 are explored in this study.

2.2. Reference: Precipitation types from Multi-Radar Multi-Sensor (MRMS) system

The MRMS system provides a suite of severe weather and Quantitative Precipitation Estimates (QPE) products by integrating operational radar observations with rain gauge observations, atmospheric model analyses, and satellite data (Zhang et al., 2016). The products are available at 0.01° (~ 1 km) and 2 min resolution across the CONUS and southern Canada. As with satellite precipitation algorithms, the MRMS QPE products use different empirical relationships for different precipitation types. An automated surface precipitation classification is employed such that appropriate relationships between radar variables and precipitation rates are derived. There are seven categories of surface precipitation types: 1) warm stratiform rain, 2) cool stratiform rain, 3) convective rain, 4) tropical–stratiform rain mix, 5) tropical–convective rain mix, 6) hail, and 7) snow. The description of each precipitation type can be referenced from Zhang et al. (2016). Kirstetter et al. (2012; 2014) set up a standardized reference for Global Precipitation Measurement (GPM) ground validation based on MRMS called GV-MRMS (Kirstetter et al. 2018). This high quality GV-MRMS precipitation type product is aggregated to a temporal scale of 30 min to mitigate uncertainty due to temporal matching and the indirect link between cloud-top observations and precipitation processes, and is used as the reference in this study.

2.3. Numerical Weather Prediction Model based environmental features

As mentioned earlier, the relationship between cloud top information from GEO sensors and surface precipitation is under-constrained (Kirstetter et al., 2018). Environmental variables from NWP models are used to complement the GEO observations and improve precipitation retrievals at the surface. To mitigate the overestimation (underestimation) of GEO retrieved rainfall in dry (wet) environments, Relative Humidity (RH) and Precipitable Water (PW) can be

used (Ba and Gruber, 2001; Kuligowski et al., 2010; Vicente et al., 1998). Recently Min et al. (2018) demonstrated the relevance of other environmental features for precipitation detection and quantification. In this study, we use NWP information to classify different precipitation processes. Using NWP model data is consistent with the reference GV-MRMS precipitation type classification that uses NWP model-based environmental variables (Zhang et al., 2016). GV-MRMS uses the Surface Temperature (ST) and Wet Bulb Temperature (WBT) to identify Snow, Cool Stratiform, and Tropical types. The zero-hour analysis of the next-generation hourly updated assimilation and model forecast cycle Rapid Refresh (RAP) is used. RAP has been part of the NOAA National Centers for Environmental Prediction (NCEP) operational suite since May 2012 (Benjamin et al., 2016). Following Grams et al. (2014), 19 features are derived from the RAP in addition to ST and WBT (Table 2). Surface-based convective available potential energy (CAPE) and lapse rates are indications of atmospheric stability which discriminate different updraft strengths in continental and maritime air masses. The vertical wind shear-based variables discriminate storm modes, such as weak shear for tropical systems and strong shear for more organized convection. Other variables provide atmospheric moisture content and degree of instability (for details the reader is referred to Grams et al. 2014). Note that radar reflectivity fields used in MRMS are assimilated in RAP and details on how assimilated reflectivity fields affect the training dataset can be found in Benjamin et al., (2016).

2.4. Data characteristics and Pre-processing

The study has been carried out across CONUS using the summer season (JJAS) of 2018. The spatial resolution of the analysis is the ABI native resolution. GV-MRMS products are spatially aggregated to match the resolution of ABI at the 30-min temporal scale. Conservative

quality controls were applied to the resampled GV-MRMS data to derive the reference. Firstly, only the most trustworthy GV-MRMS data are identified with the high threshold for Radar Quality Index (RQI). A threshold of 98% is used for all precipitation types except for Hail. Given that hail has low occurrence and its detection is comparatively less impacted by radar sampling than other precipitation types, a lower threshold of 90% is used to get sufficient samples. Secondly, to target homogeneous precipitation types in the matched ABI pixel, only resampled GV-MRMS pixels with at least 98% of the same precipitation type are selected with the exception of 90% for Convective and Tropical Convective Mix, 80% for Hail. The remaining “Mix” class pixels are not used. Because precipitation intermittency impacts the satellite signal (Upadhyaya et al., 2020), reference grids including partial precipitation are further removed, and only grids with percent rain greater than 95% and less than 5% are used for analysis.

The size of the dataset and the distribution of precipitation types across months is summarized in Table 3 after applying the quality controls. Note that due to the low sample size of GV-MRMS-GV precipitation type “Snow” in summertime, it has not been considered in this study. The first 70% of data each month is used for training and the remaining 30% is used for testing. To avoid information leakage, training and testing cases are checked to be distinct and have uniform spatial and temporal representation.

Initial experiments (not shown here) suggest that using imbalanced data for ML training leads to biased classification. Following the recommendations by Sun et al. (2009), balanced training data were generated using random sampling techniques. The balanced training and the entire validation sample sizes are shown in Table 3.

2.5. Random Forest and evaluation statistics

Accepted Article

A random forest (RF) machine learning technique is used to classify precipitation types. RF is shown to be a robust approach for classification and has been applied to diverse applications including in precipitation science (Kuhnlein et al., 2014; Lazri and Ameer, 2018; Ouallouche et al., 2018). As the name suggests “Random Forest” is basically a “Forest” of decision trees building on two concepts of “Randomness” (Breiman, 2001). First, the training data is randomly sampled following the concept of “Bagging” (Breiman, 1996). Second, a subset of randomly selected features are used to build each decision tree. Díaz-Uriarte et al. (2006) summarizes several advantages of RF: (1) RFs produce unbiased estimates with low variance, (2) there is minimal overfitting; (3) RFs are effective in building non-linear relationships between the predictors and the predictand; and (4) there are only a few parameters that require tuning.

One RF parameter is the number of bootstrap samples (n) to develop n number of trees. For each RF tree, m is the number of randomly selected features. We used 500 trees (n) with $m = \sqrt{(no. of features)} = 16$. A sensitivity analysis was performed to fine-tune these parameters, but did not show significant impact on the classification accuracy. Thus, for all experiments these two parameters were kept as indicated above. The ML framework scikit-learn in python (Pedregosa et al., 2011) is used for implementing random forest.

To evaluate the model, a confusion matrix is built with the probability distributions of RF classified precipitation types. It is used along with individual type (classified precipitation type) accuracy and overall classification accuracy. The statistics derived from this matrix are briefly described below. More details can be found in Congalton (1991).

Overall Accuracy: the proportion of total grids correctly classified irrespective of individual class accuracy. It is computed as the ratio (in percent) of the number of correctly classified data points to the total number of data points.

Kappa Coefficient: the difference between the obtained agreement and the agreement expected by chance. A Kappa value of 0.85 indicates that there is 85% better agreement than by chance alone. Typical values for Kappa are -1 (no agreement), 0 (random agreement), and 1 (perfect agreement).

Class Accuracy: accuracy obtained for each class. It is calculated as the ratio of the number of correctly identified data points and the reference sample size for a given class expressed in percent. It is equivalent to probability of detection (POD) for each class. These values are the diagonal elements of the contingency matrix (Fig. 1)

Additional details on the characteristics of RF and accuracy assessment are provided as required in Section 3 sections.

3. Results and Discussion

3.1. Model accuracy assessment

In case of multi-class classification problems, RF inherently builds different forests/models for each class. As an output, RF computes the probability that any given sample belongs to each class (precipitation type), and the sample is assigned to the precipitation type associated with the highest probability. While the developed model is probabilistic in nature, in general it is used in a deterministic manner. In this study, we analyze results both from the deterministic and probabilistic perspectives.

In terms of deterministic classification, Fig. 1 shows the normalized contingency matrix, along with other statistics reported in Section 2.5. It is normalized across all GV-MRMS precipitation types for ease of inter-comparison between different types. It is complemented by Fig. 2 showing each panel of the confusion matrix (i.e. each combination of RF classified and

reference precipitation types) as two frequency curves of probabilities estimated by the RF model: the probability of the RF classified type (pink curve) and the probability of the GV-MRMS reference precipitation type (green curve). Note that diagonal cells have only one pink curve because the samples are correctly classified. The number at the top of each panel represents normalized contingency matrix elements (from Fig. 1) along with sample size. This graph indicates whether the probabilities of the RF classified type and reference class type are close (in other words they are misclassified with a small difference in probabilities), or otherwise.

Regarding No-Precipitation, (used as NoPrecip in Figures 1 and 2) the RF model has an accuracy of 93% with 5% misclassified in warm stratiform (Figure 1). False alarms from other classes are low (Vertical column of No-Precipitation) with 2% of warm stratiform and 1% of convective events misclassified as No-Precipitation. From Fig. 2, one can observe that the estimated probability is very high i.e. curve peak close to 1, indicating that the rain and no-rain separation is robust with the RF model. Distributions associated with Misses (NoPrecip reference cases misclassified into other precipitation types) display low overlap between the reference distribution and the distribution associated with the other precipitation type. Note that the overlap is slightly higher in case of NoPrecip misclassified into warm precipitation, consistent with the 5% reported in Fig. 1. The challenge will be to classify precipitating pixels to the appropriate class.

Stratiform Types include Cool Stratiform (CoolStrat), Warm Stratiform (WarmStrat) and Tropical Stratiform/Mix (Trp_StratMix). From Figure 1, the CoolStrat type has accuracy of 97%, with 3% miss-classification in another stratiform category i.e. Warm Stratiform. This high accuracy can be attributed to the use of the Wet Bulb Temperature as a RF predictor feature while MRMS uses it also to identify CoolStrat. CoolStrat false alarms are also low, again with 1% of

Accepted Article

warm-stratiform pixels incorrectly classified as cool stratiform. Similar to No Precipitation, the robustness of the RF model is very high, i.e.: the probability curve is close to 1 (Figure 2). Warm stratiform has an accuracy of about 72% and is mostly misclassified as Tropical Stratiform Mix (11%) and Convective (10%) types. Warm Stratiform displays large false alarms also from the same types (around 20% from each). From Fig. 2, one can see that there are significant overlaps in the probability curves for Trp_StratMix and Convective in Warm Stratiform. This indicates that the lower probabilities estimated by the RF for the estimated and observed type are close so these types may generate misclassification. A more detailed analysis on these effects will be reported in the companion article (Upadhyaya et al., 2021). The Tropical Stratiform/Mix (TSM) has an accuracy of 64%. It is misclassified as Tropical Convective/Mix (TCM;14%), in Warm stratiform (20%), and as Convective (4%). False alarms in this class come from the same classes: TCM (around 26%), warm stratiform (11%) and Convective (9%).

Convective Types include Hail, Convective, and Tropical Convective/Mix. The RF model has accuracy of 55% for TCM. A majority of misses are classified instead as TSM (26%) followed by in the Convective type (10%) as confirmed by the considerable overlap in RF estimated probability curves (Fig. 2). TCM has also significant false alarms from the same two types (around 12% from each class). The Convective type has accuracy of only 47% and it is mis-classified as all other rain classes except Cool Stratiform and No-Precipitation. As expected, there is considerable overlap between the probability curves, confirming that the RF model is challenged in separating this class from others. In the companion article (Upadhyaya et al., 2021) the associated reasons are explored. Hail has a higher detection score (70%) than the other convective types. Most of its misclassification occurs in the convective class (26%) associated with large overlap in the probability curves. Most false alarms are also coming from the Convective type.

Accepted Article

The developed prognostic model shows skillful performance in identifying the occurrence/non-occurrence of precipitation (accuracy of 97%; Kappa coefficient of 0.9) and precipitation processes, with an overall classification accuracy of 76% and Kappa coefficient of 0.56. The precipitation types No-Precipitation and Cool-Stratiform are classified with very high classification accuracy and probability (close to 1). The classification accuracy for types Warm Stratiform and Hail are relatively high, but the probability curves peak at values lower than 0.5 for the remaining types, indicating decreased robustness in the RF estimation for these types. As expected, frequently misclassified types (i.e., contingency matrix cells with large percentages) tend to be associated with larger amounts of overlap between the two probability curves than for the types that are misclassified less frequently. As indicated by stacked bars and highlighted by cells from the contingency matrix in Fig. 1, one can note a shifting loop in the misclassification trend from Convective to Tropical Convective/Mix to Tropical Stratiform/Mix to Warm Stratiform and back to Convective, and between Hail and Convective types. While these precipitation types reflect different precipitation processes, it is challenging to separate them deterministically from the GOES16 ABI observations. This motivates the use of a probabilistic rather than a deterministic classification to increase the information content used in the interpretation of identified precipitation types and in rate estimates.

3.2. Case Studies

Two randomly selected case events are evaluated visually and quantitatively by analyzing the contingency matrix in Figure 2. The RF classification results are compared with GV-MRMS and with SCaMPR (Kuligowski et al., 2010, 2016).

Accepted Article

The first event on 6 July 2018 at 0900 UTC is a case of a well-developed summer precipitation system in the southern Great Plains (Fig. 3). Figures 3a and 3d show the reference GV-MRMS precipitation types and the RF-estimated precipitation types, respectively. The RF probabilities estimated for each precipitation type are given as supplementary Figure S1, and the probability of the estimated dominant type at each pixel is shown in Figure 3f. Comparing Figs. 3a and 3d, it can be observed that the RF model captures the occurrence of precipitation and the spatial shape of the precipitation event, yet there is noticeable overestimation of the raining area. Most of the precipitation false alarm area is classified as warm stratiform, which is consistent with the findings in Section 3.1. Sources of error causing this misclassification may include sub-pixel rainfall variability along the edges of rainy areas associated with the satellite sampling resolution (i.e., non-uniform beam filling (NUBF) as reported in Kirstetter et al., 2012, 2013 and Upadhyaya et al., 2020). Other sources of error can arise from the spatio-temporal matching between ABI and GV-MRMS, and possibly also internal MRMS procedures to avoid virga (Zhang et al., 2016). In contrast, SCaMPR misses large areas of precipitation (Figure 3c).

Heavy precipitation regions (Figure 3b) such as the narrow stream in the north side and the small patch of heavy rain rates in the southwest that are classified as Convective and Hail types by GV-MRMS (Figure 3a) are well-captured by the RF model estimates (Figure 3d). However, the spatial extent of the Convective and Hail areas are overestimated by RF estimates relative to the GV-MRMS precipitation types.

Note that precipitation types from GV-MRMS have their own sources of uncertainties and limitations, and might not optimally transfer to the satellite classification. Using the full information content of the RF retrievals through probabilities of precipitation types makes more sense than attempting to deterministically and exactly replicate the reference. From the dominant

probability map (Fig. 3f) and Figure S1, and from the contingency matrix (Fig. 4), it can be observed that regions of misclassification are generally associated with lower dominant RF probability values. The estimated probabilities for both the misclassified and the correct reference types are less than 0.5, indicating that the RF estimates in these regions are uncertain. Such regions where the probability is less than 0.5 are highlighted as grey in Fig. 3e, and labelled as “uncertain” type. To account for the detection of precipitation, this uncertain type is further refined into Uncertain Rain (U:Rain) and Uncertain No-Rain (U:NoRain).

Several different sets of rules can be developed to identify uncertain regions and types U:Rain and U:NoRain. Three of them are displayed and compared in Fig. 5:

1. “Uncertain” can be defined such that the dominant estimated probability is less than 0.6, and this is subdivided into U:NoRain if the No-Precipitation type probability is in the range [0.4 - 0.6] or U:Rain if the No-Precipitation type probability is less than 0.4 (Fig. 5d); or
2. “Uncertain” can be defined as when the first dominant estimated probability is less than twice as high as the second dominant probability; if either this first dominant or the second dominant type is No-Precipitation then it is classified as U:NoRain (Fig. 5e); or
3. A threshold of 0.5 is used to define the dominant probability; to separate U:NoRain, the same rule as in Rule 2 is applied (Fig. 5f).

By visually comparing the three maps, the Rule 3 map appears to be closest to GV-MRMS. It is evident that these sets of deterministic rules include some subjectivity and may not necessarily generalize to other places and times. Further, in GV-MRMS some regions at the edges of the rain field can be highlighted as “Uncertain” (Fig. 5c) because the dominant precipitation type is less than 80% in the space and time scale of aggregation. Note that the GV-MRMS mask associated

with RQI is not visually used since it would remove significant precipitation areas (Pink patches in Fig. 5b) indicating there may be some uncertainty in GV-MRMS at lower RQI.

The second randomly selected event occurred on 12 August 2018 at 2300 UTC (Fig. 6) across complex terrain of the western CONUS. It is characterized by scattered convection (also called air mass thunderstorms). As in the previous case the rain area is overestimated by the classification scheme. From Fig. 6c it can be noted that SCaMPR picks up the larger systems located in the west but misses several small scattered events, while the RF estimations correctly detect almost all scattered convective cells. From the contingency matrix and probability curves (Fig. 7), all precipitation types are estimated with low probability except no-precipitation. Some Hail regions and No-precipitation pixels are estimated with high probability, but large rain areas are estimated as uncertain types (Figs. 6e-f).

RQI values are generally low across the western CONUS which indicates that the MRMS coverage is limited (Fig. 6a). This case highlights the potential of GEO sensor observations for complementing the limited radar coverage across the Intermountain West.

3.3. Feature Importance and optimal model selection

As a diagnostic tool, the Random Forest model provides feature importance from training. The feature importance indicates how much each feature contributes to decreasing the mean Gini impurity (a measure of misclassification) in a forest (Pedregosa et al., 2011). This measure is normalized to sum to 1 and is ranked accordingly. The feature importance can be used to analyze which features are important for the classification and to select features in order to create more interpretable and less computationally expensive models if the application requires it.

Figure 8 shows the first most important features contributing up to 65% of total feature importance. Most of the RAP model-based environmental features have higher importance than satellite features, especially the temperature-related RAP features. Regarding satellite features, D-BTD and textures derived from D-BTD have higher importance than other satellite features. The interested reader is referred to the companion article (Upadhyaya et al., 2021) of the article that details which features are contributing to each precipitation type classification.

In this study, a total of 260 features are used, most of which may contribute little to the accuracy of classification. A feature reduction experiment is carried out to select a more parsimonious model; i.e., with the fewest possible features without significantly compromising the classification accuracy. Figure 9 illustrates the reduction classification accuracy for each precipitation type from gradually removing features in reverse order of importance (based on the cumulative feature importance values). The threshold used for cumulative feature importance along with the number of features are given in the x-axis in Fig. 9. From Fig. 9 and Fig. 1, recall that the accuracy of each precipitation type is different with Cool Stratiform and No-Precipitation with highest accuracy (>90%), followed by Warm Stratiform and Hail (>70%) , Tropical classes (<70%) and Convective (<50%).

By removing the last 1% of features based on cumulative importance, the number of features is reduced from 260 to 247 without a significant decrease in classification accuracy for all precipitation types. Further removal of features noticeably reduced the accuracy in the convective precipitation type category (i.e., Hail, Convective and Tropical Convective/Mix), but the overall accuracy remains almost identical since these types represent a relatively small fraction of the total number of pixels. The overall accuracy statistic is most sensitive to the accuracy of No-Precipitation and Warm stratiform types because their proportions are significantly higher than

other types. This highlights that the individual class accuracies and contingency matrices should be used instead of the overall accuracy to assess classification models so that effects on individual classes are not missed. A drop-in accuracy can be noted when the number of features is reduced from 63 to 52 (i.e., the cumulative importance threshold is reduced from 65% to 60%). It indicates that important features for the Convective types are removed which are the last few features in Fig. 8. These include D-BTDs, BTDs between WV and IR and single channel features such as T6.2, T6.2 Contrast, T11.2 mean, T12.3. The companion article (Upadhyaya et al., 2021) focuses specifically on identifying and interpreting important features for each precipitation type.

4. Conclusions

Using the Ground Validation Multi-Radar/Multi-Sensor (GV-MRMS) system as a benchmark, a prognostic machine learning-based precipitation type classification model is developed by deriving a comprehensive set of features from ABI observations and numerical weather prediction data. While the developed ML model using Random Forests (RF) is probabilistic in nature, in general precipitation type classifications are used in a deterministic manner. A detailed analysis comparing the deterministic and probabilistic perspectives is performed and leads to key highlights:

1. The overall accuracy of RF classification is 75.9% and the Kappa coefficient is 0.56;
2. Precipitation types No-Precipitation and Cool-Stratiform are classified with very high classification accuracy (>90%) and probability (close to 1);
3. The classification accuracy for types Warm Stratiform and Hail are relatively high (>70%).

However, the corresponding probability curves peak at values lower than 0.5, indicating decreased robustness in the RF estimation for these types;

- Accepted Article
4. Challenges exist in identifying the Convective and Tropical types;
 5. There is a shifting loop in terms of misclassification from Convective to Tropical Convective/Mix to Tropical Stratiform/Mix to Warm Stratiform and back to Convective, and between Hail and Convective types. It brings into question the relevance of a deterministic classification, and motivates the use of a probabilistic classification to improve precipitation retrievals.
 6. A feature reduction experiment showed that the number of model predictors can be reduced from 260 to 63 (i.e., a model with the fewest possible features) without significantly compromising the classification accuracy.

Overall, Part I of this article focused on prognostic modeling; i.e., the design, training, and assessment of a machine learning-based model for precipitation type and processes classification. Part II of the article focuses on the interpretation of the ML model with the objective to identify predictors relevant to different precipitation processes.

Acknowledgements

We are very much indebted to the teams responsible for the GOES-R, MRMS and SCaMPR products. Datasets for this research are available in these in-text data citation references: Kuligowski et al. (2016), Kirstetter et al. (2012; 2014), Benjamin et al. (2016). Authors acknowledge Y. Derin for helping with downloading and extracting environmental predictors. We also thank the editor and reviewers for their suggestions and comments that helped to improve the manuscript. Funding for this research was provided by the GOES-R Series Risk Reduction program, which provided support to the Cooperative Institute for Mesoscale Meteorological

Studies at the University of Oklahoma under Grant NA16OAR4320115. P. Kirstetter acknowledges support from NASA Global Precipitation Measurement Ground Validation program under Grant NNX16AL23G and Precipitation Measurement Missions program under Grant 80NSSC19K0681. The contents of this paper are solely the opinions of the authors and do not constitute a statement of policy, decision, or position on behalf of NOAA or the U.S. Government.

References

- Ba, M. B., & Gruber, A. (2001). GOES multispectral rainfall algorithm (GMSRA). *Journal of Applied Meteorology*, 40(8), 1500-1514.
- Baum, B. A., & Platnick, S. (2006). Introduction to MODIS cloud products. In *Earth science satellite remote sensing* (pp. 74-91). Springer, Berlin, Heidelberg.
- Benjamin, S. G., Weygandt, S. S., Brown, J. M., Hu, M., Alexander, C. R., Smirnova, T. G., ... & Lin, H. (2016). A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Monthly Weather Review*, 144(4), 1669-1694.
- Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123-140.
- Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote sensing of environment*, 37(1), 35-46.
- Díaz-Uriarte, R., & De Andres, S. A. (2006). Gene selection and classification of microarray data using random forest. *BMC bioinformatics*, 7(1), 3.
- Giannakos, A., & Feidas, H. (2013). Classification of convective and stratiform rain based on the spectral and textural features of Meteosat Second Generation infrared data. *Theoretical and applied climatology*, 113(3-4), 495-510.
- Grams, H. M., Kirstetter, P. E., & Gourley, J. J. (2016). Naïve Bayesian precipitation type retrieval from satellite using a cloud-top and ground-radar matched climatology. *Journal of Hydrometeorology*, 17(10), 2649-2665.
- Grams, H. M., Zhang, J., & Elmore, K. L. (2014). Automated identification of enhanced rainfall rates using the near-storm environment for radar precipitation estimates. *Journal of Hydrometeorology*, 15(3), 1238-1254.

- Hall-Beyer, M. (2000). GLCM texture: a tutorial. *National Council on Geographic Information and Analysis Remote Sensing Core Curriculum*, 3.
- Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6), 610-621.
- Hong, Y., Hsu, K. L., Sorooshian, S., & Gao, X. (2004). Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *Journal of Applied Meteorology*, 43(12), 1834-1853.
- Kirstetter, P. E., Hong, Y., Gourley, J. J., Cao, Q., Schwaller, M., & Petersen, W. (2014). Research framework to bridge from the Global Precipitation Measurement Mission core satellite to the constellation sensors using ground-radar-based national mosaic QPE. *Remote sensing of the terrestrial water cycle*, 61-79.
- Kirstetter, P. E., Hong, Y., Gourley, J. J., Chen, S., Flamig, Z., Zhang, J., ... & Amitai, E. (2012). Toward a framework for systematic error modeling of spaceborne precipitation radar with NOAA/NSSL ground radar-based National Mosaic QPE. *Journal of Hydrometeorology*, 13(4), 1285-1300.
- Kirstetter, P. E., Karbalaee, N., Hsu, K., & Hong, Y. (2018). Probabilistic precipitation rate estimates with space-based infrared sensors. *Quarterly Journal of the Royal Meteorological Society*, 144, 191-205.
- Kühnlein, M., Appelhans, T., Thies, B., & Nauß, T. (2014). Precipitation estimates from MSG SEVIRI daytime, nighttime, and twilight data with random forests. *Journal of Applied Meteorology and Climatology*, 53(11), 2457-2480.

- Accepted Article
- Kuligowski, R. J. (2002). A self-calibrating real-time GOES rainfall algorithm for short-term rainfall estimates. *Journal of Hydrometeorology*, 3(2), 112-130.
- Kuligowski, R. J., Li, Y., Hao, Y., & Zhang, Y. (2016). Improvements to the GOES-R rainfall rate algorithm. *Journal of Hydrometeorology*, 17(6), 1693-1704.
- Lazri, M., & Ameer, S. (2018). Combination of support vector machine, artificial neural network and random forest for improving the classification of convective and stratiform rain using spectral features of SEVIRI data. *Atmospheric research*, 203, 118-129.
- Le, M., & Chandrasekar, V. (2012). Precipitation type classification method for dual-frequency precipitation radar (DPR) onboard the GPM. *IEEE transactions on geoscience and remote sensing*, 51(3), 1784-1790.
- Le, M., & Chandrasekar, V. (2021). Graupel and hail identification algorithm for the dual-frequency precipitation radar (DPR) on the GPM core satellite. *Journal of the Meteorological Society of Japan*. Ser. II.
- Lovejoy, S., & Austin, G. L. (1979). The delineation of rain areas from visible and IR satellite data for GATE and mid-latitudes. *Atmosphere-ocean*, 17(1), 77-92.
- Min, M., Bai, C., Guo, J., Sun, F., Liu, C., Wang, F., ... & Dong, L. (2018). Estimating summertime precipitation from Himawari-8 and global forecast system based on machine learning. *IEEE Transactions on Geoscience and Remote Sensing*, 57(5), 2557-2570.
- Mohanaiah, P., Sathyanarayana, P., & GuruKumar, L. (2013). Image texture feature extraction using GLCM approach. *International journal of scientific and research publications*, 3(5), 1.

- Ouallouche, F., Lazri, M., & Ameer, S. (2018). Improvement of rainfall estimation from MSG data using Random Forests classification and regression. *Atmospheric Research*, 211, 62-72.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
- Petković, V., Orescanin, M., Kirstetter, P., Kummerow, C., & Ferraro, R. (2019). Enhancing PMW satellite precipitation estimation: Detecting convective class. *Journal of Atmospheric and Oceanic Technology*, 36(12), 2349-2363.
- Rampun, A., Strange, H., & Zwiggelaar, R. (2013, June). Texture segmentation using different orientations of GLCM features. In *Proceedings of the 6th International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications* (pp. 1-8).
- Schmit, T. J., Griffith, P., Gunshor, M. M., Daniels, J. M., Goodman, S. J., & Lebar, W. J. (2017). A closer look at the ABI on the GOES-R series. *Bulletin of the American Meteorological Society*, 98(4), 681-698.
- So, D., & Shin, D. B. (2018). Classification of precipitating clouds using satellite infrared observations and its implications for rainfall estimation. *Quarterly Journal of the Royal Meteorological Society*, 144, 133-144.
- Sun, Y., Wong, A. K., & Kamel, M. S. (2009). Classification of imbalanced data: A review. *International journal of pattern recognition and artificial intelligence*, 23(04), 687-719.
- Tebbi, M. A., & Haddad, B. (2016). Artificial intelligence systems for rainy areas detection and convective cells' delineation for the south shore of Mediterranean Sea during day and nighttime using MSG satellite images. *Atmospheric research*, 178, 380-392.

- Thies, B., Nauß, T., & Bendix, J. (2008). Precipitation process and rainfall intensity differentiation using Meteosat second generation spinning enhanced visible and infrared imager data. *Journal of Geophysical Research: Atmospheres*, 113(D23).
- Tian, B., Shaikh, M. A., Azimi-Sadjadi, M. R., Haar, T. H. V., & Reinke, D. L. (1999). A study of cloud classification with neural networks using spectral and textural features. *IEEE transactions on neural networks*, 10(1), 138-151.
- Tjemkes, S. A., Van de Berg, L., & Schmetz, J. (1997). Warm water vapour pixels over high clouds as observed by Meteosat. *Beitrage zur Physik der Atmosphäre-Contributions to Atmospheric Physics*, 70(1), 15-22.
- Upadhyaya, S. A., Kirstetter, P. E., Gourley, J. J., & Kuligowski, R. J. (2020). On the propagation of satellite precipitation estimation errors: from passive microwave to infrared estimates. *Journal of Hydrometeorology*, 21(6), 1367-1381.
- Upadhyaya, S. A., Kirstetter, P. E., Kuligowski, R. J., & M. Searls (2021). Classifying precipitation from GEO Satellite Observations: Diagnostic Model. *Journal of Geophysical Research: Atmosphere* (Submitted).
- Upadhyaya, S., & Ramsankaran, RAAJ. (2014). Multi-index rain detection: a new approach for regional rain area detection from remotely sensed data. *Journal of Hydrometeorology*, 15(6), 2314-2330.
- Vicente, G. A., Scofield, R. A., & Menzel, W. P. (1998). The operational GOES infrared rainfall estimation technique. *Bulletin of the American Meteorological Society*, 79(9), 1883-1898.

Xian, G. M. (2010). An identification method of malignant and benign liver tumors from ultrasonography based on GLCM texture features and fuzzy SVM. *Expert Systems with Applications*, 37(10), 6737-6741.

Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., ... & Arthur, A. (2016). Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bulletin of the American Meteorological Society*, 97(4), 621-638.

Zvloff, A, (2020). glcm: Calculate Textures from Grey-Level Co-Occurrence Matrices. (GLCMs). R package version 1.6.5. <https://CRAN.R-project.org/package=glcm>

List of Tables:

Table 1: Categories and number of features used in study

<i>Category</i>	<i>Feature Type</i>	<i>Number of features</i>	<i>Example</i>
1	BT (Brightness Temperature)	5	T6.2
2	BTD (Brightness Temperature Difference)	10	T6.2 – T7.3
3	D-BTD (Difference of BTDs)	25	(T6.2 – T7.3) – (T8.5 – T11.2)
4	Te (GLCM Textures)	5 Textures x (5 Category 1+ 10 Category 2+ 25 Category 3) = 200	T6.2 mean
5	Ze (Satellite Zenith Angle)	1	Ze
6	Environmental features (NWP)	19	Details in Table 2

*T6.2 is read as brightness temperature of ABI channel 6.2μm

Table 2: RAP model-based environmental features used in study

<i>Sl No</i>	<i>Environmental Variable</i>
1	Vertically integrated precipitable water (kg/m2)
2	1000-700-hPa mean relative humidity (%)
3	900-hPa relative humidity (%)
4	850-hPa relative humidity (%)
5	700-hPa relative humidity (%)
6	500-hPa relative humidity (%)
7	Surface equivalent potential temperature (K)
8	Surface-based convective available potential energy (CAPE) (J/kg)
9	Surface temperature (C)
10	850-hPa temperature (K)
11	700-hPa temperature (K)
12	500-hPa temperature (K)
13	Height of 0C isotherm (km)
14	Wind shear from surface to 850 hPa (m/s)
15	Wind shear from surface to 700 hPa (m/s)
16	Wind shear from surface to 500 hPa (m/s)
17	850-500-hPa lapse rate (K/km)
18	850-700-hPa lapse rate (K/km)
19	Wet Bulb Temperature

* Note: The **bold rows** are features derived from RAP output and the other features are directly available from RAP output

List of Figures:

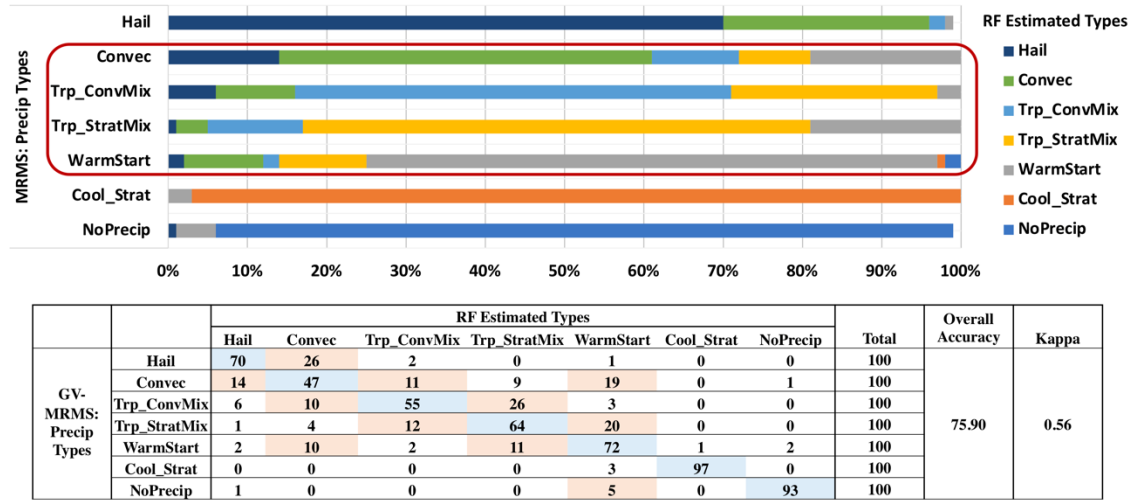


Figure 1. Normalized Contingency Matrix (in percent) as stacked bar chart with data table (lower panel) along with overall classification accuracy and Kappa coefficient. Blue highlighted cells are class accuracy values (Probability of detection) and orange highlighted cells are classes with large misclassification.

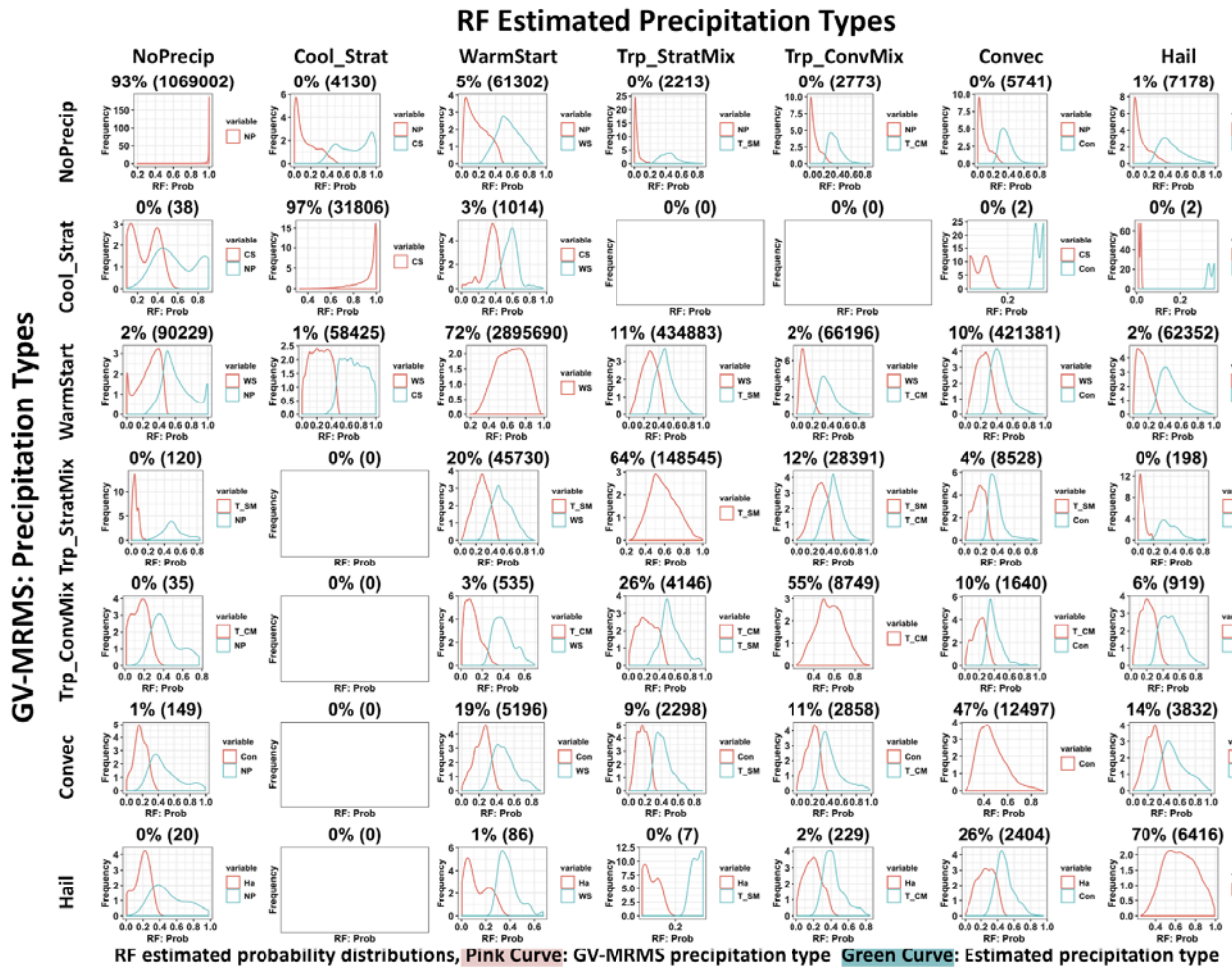


Figure 2. Normalized contingency matrix (with sample size) and frequency distribution of probabilities estimated by random forest (RF) model in terms of probabilities estimated for the reference precipitation type (pink curves) and probabilities estimated for the considered precipitation type (green curves; deterministic manner). The ideal scenario is when two curves do not overlap.

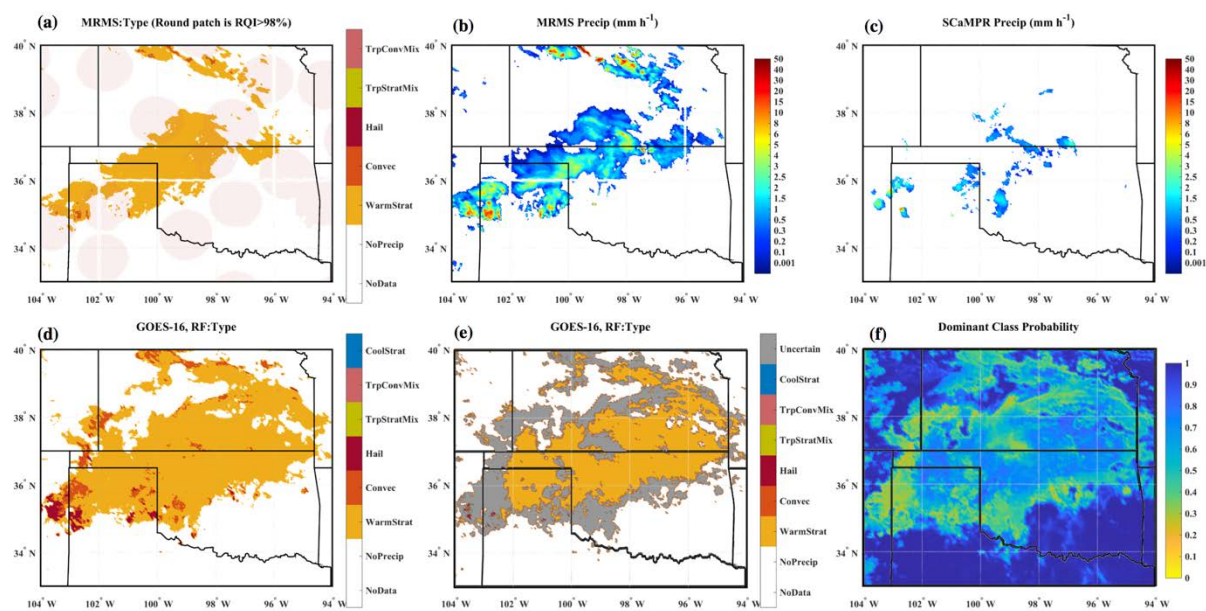


Figure 3: Case event on 26 July 2018 at 0900 UTC of (a) GV-MRMS precipitation types, (b) GV-MRMS precipitation rates, (c) SCaMPR precipitation rates, (d) estimated precipitation types, (e) estimated types highlighting uncertain regions, and (f) estimated probabilities of dominant precipitation types.

Note: White strips in Panel a and b is result from resampling GV-MRMS to the ABI scale at the boundaries of UTM zones

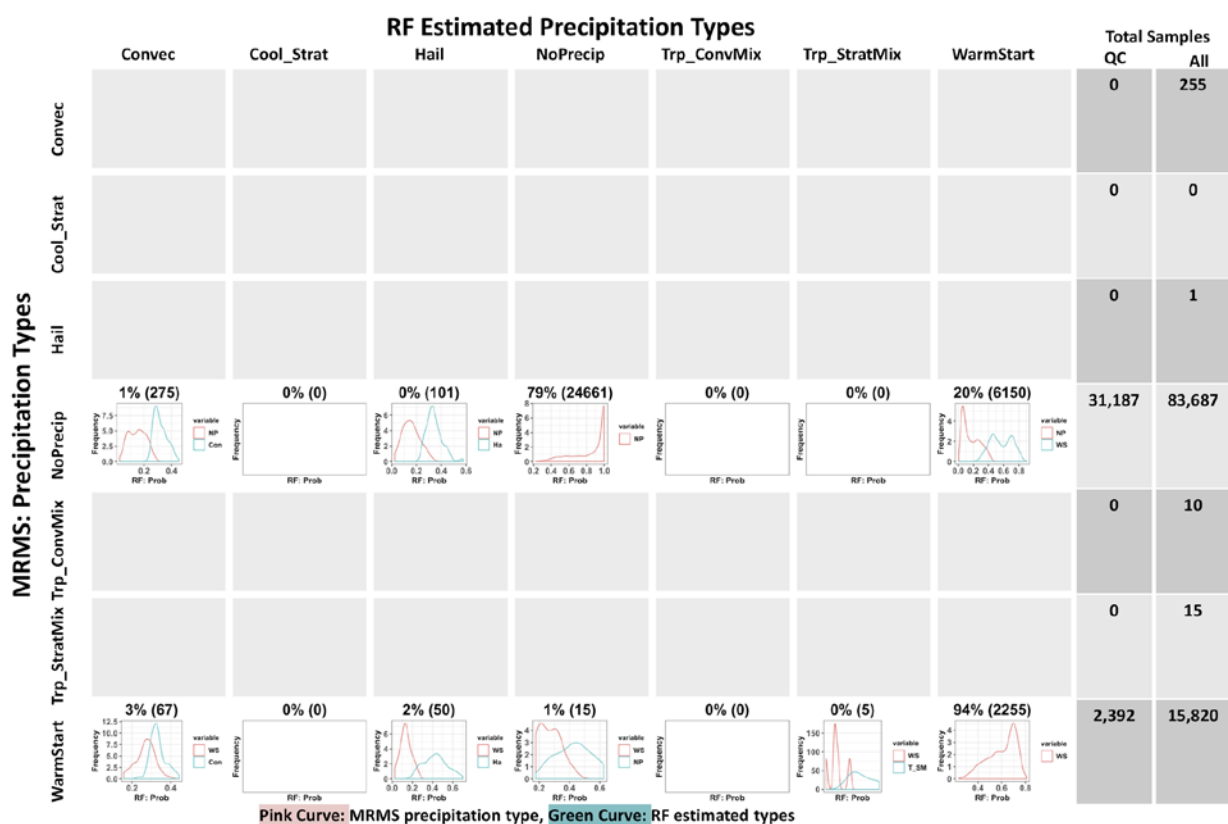


Figure 4. As in Figure 2 for the case event of July 26, 2018 at 0900 UTC using only quality control (QC) GV-MRMS observations. The last two columns indicate the total sample sizes in each GV-MRMS precipitation type with and without QC.

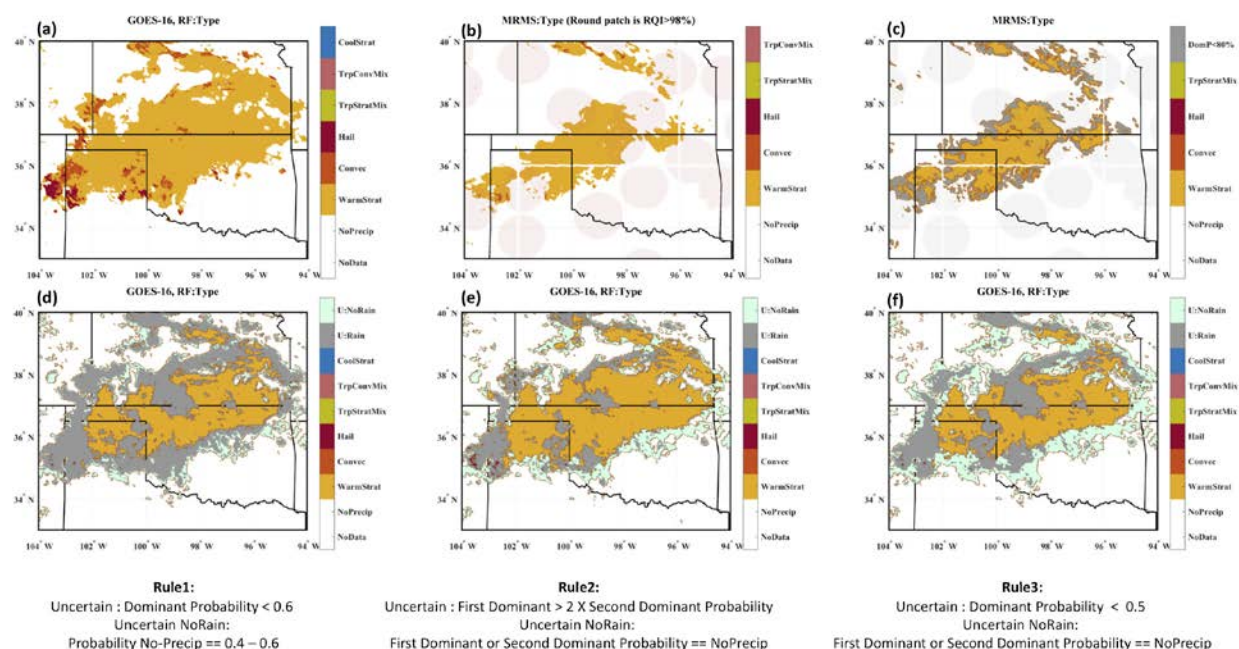


Figure 5: Case event 26 July 2018 at 0900 UTC of (a) RF dominant estimated precipitation types, (b) GV-MRMS precipitation types, (c) GV-MRMS precipitation types highlighting less certain precipitation types areas (i.e. less than 80% occurrence), (d,e,f) uncertain Rain/No-Rain areas identified with three different set of rules applied on the RF estimated precipitation types.

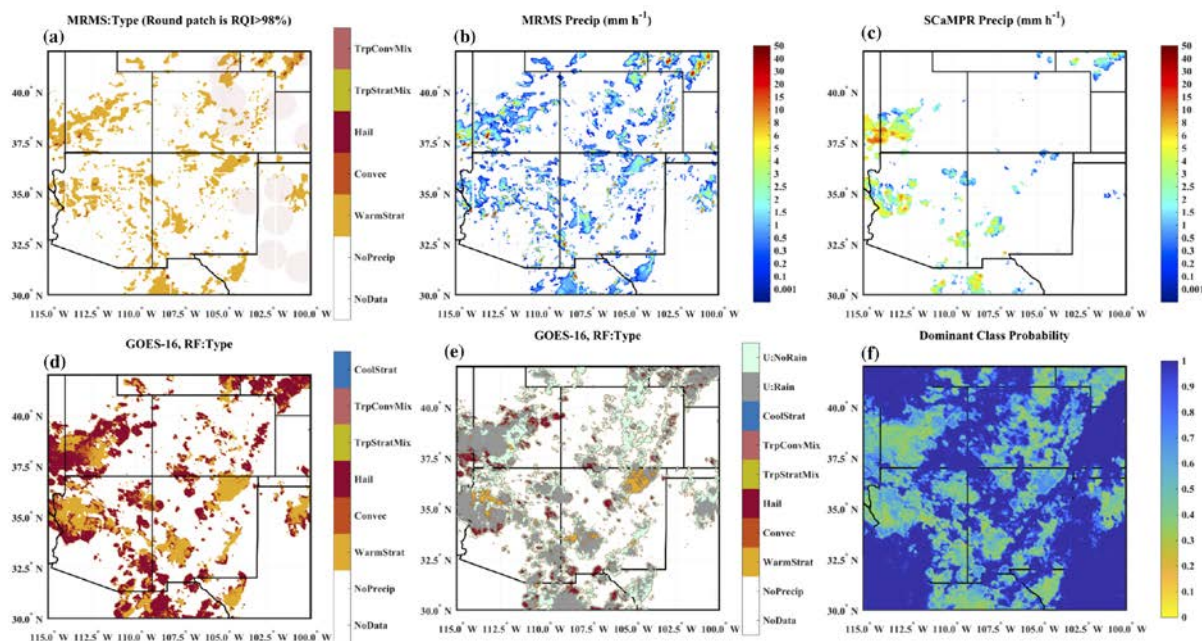


Figure 6: Case event 12 August 2018 at 2300 UTC of (a) GV-MRMS precipitation types, (b) GV-MRMS precipitation rates, (c) SCaMPR precipitation rates, (d) estimated precipitation types, (e) estimated types highlighting uncertain regions, and (f) estimated probabilities of dominant precipitation types.

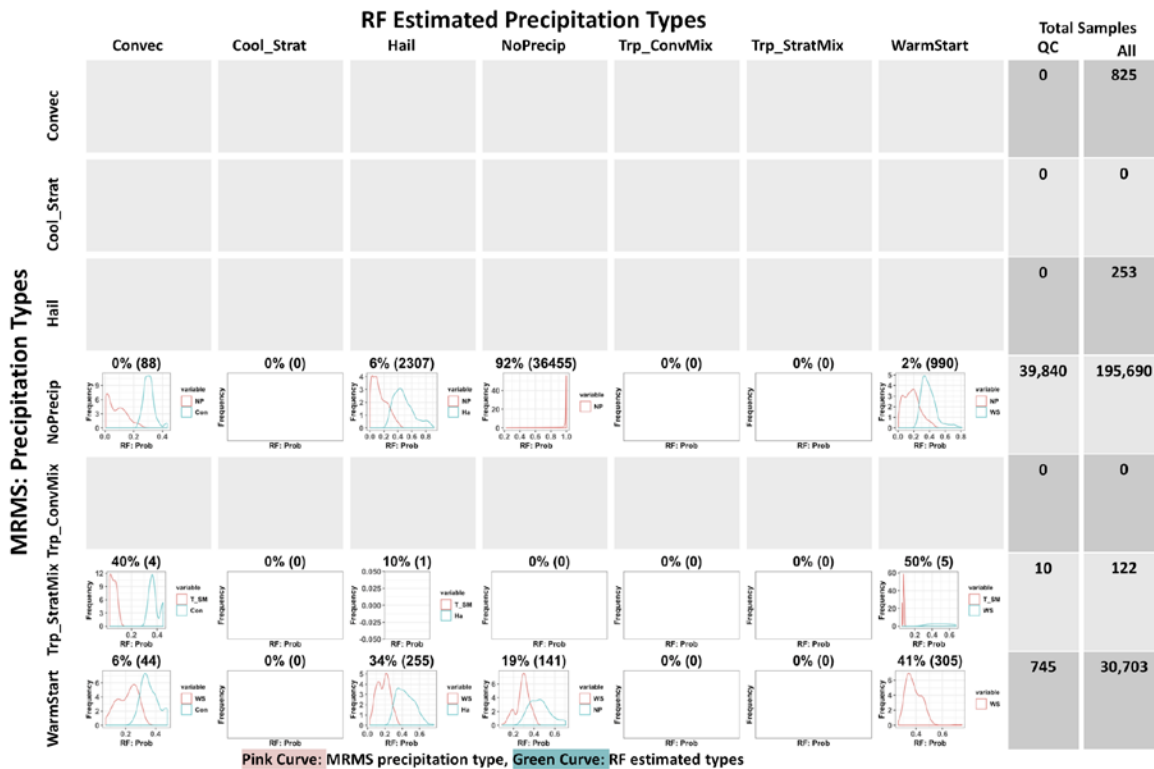


Figure 7. Contingency matrix and predicted probability graphs as in Figure 2 but for 12 August 2018 2300 UTC calculated using only quality-controlled (QC) GV-MRMS observations. Last two columns are the total sample size in each GV-MRMS precipitation type with and without QC.

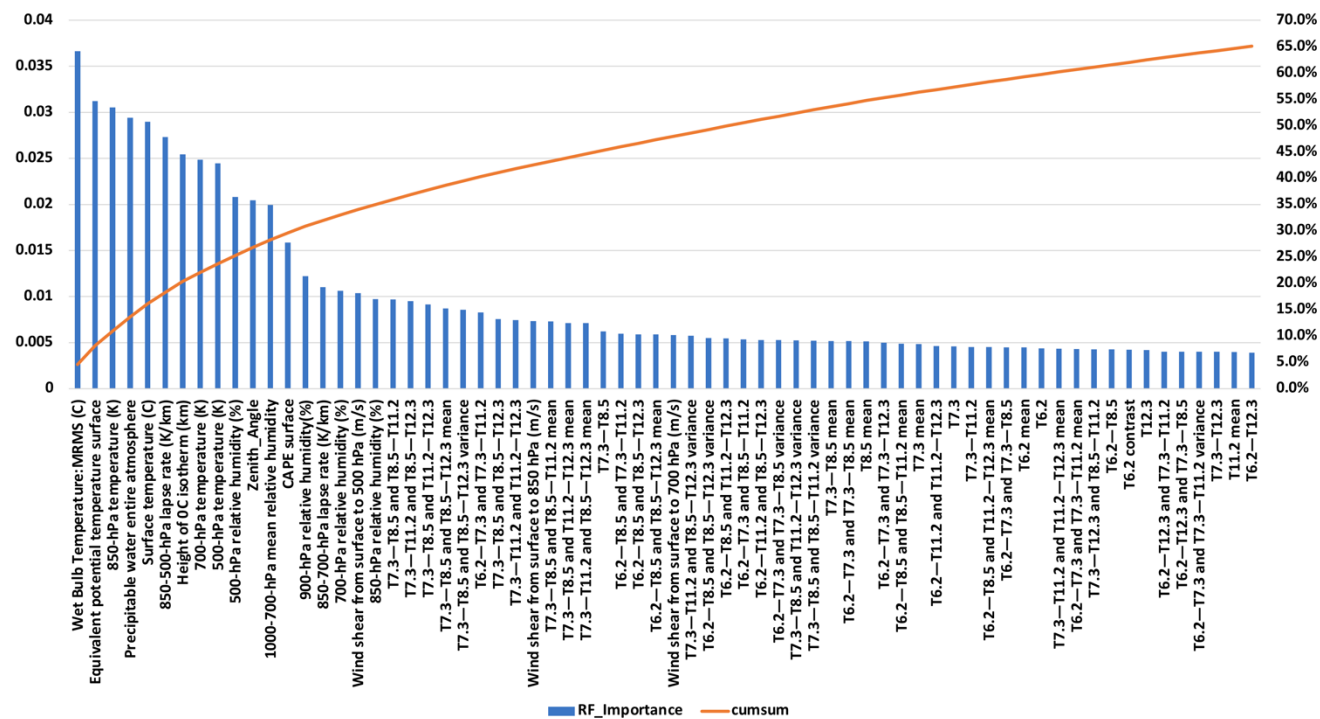


Figure 8. Most important features (contributing to 65% of the total feature importance), and their feature importance by Random Forest model in decreasing order. The cumulative importance is indicated by the orange line.

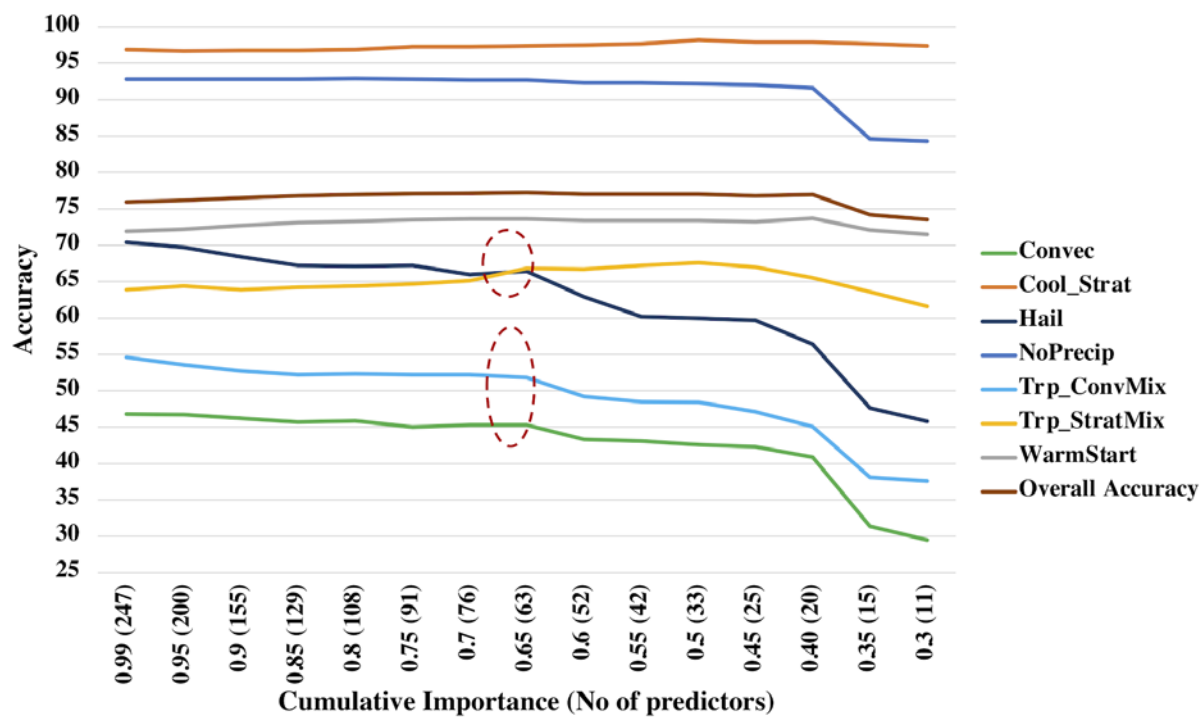


Figure 9. Feature selection/reduction based on RF feature importance